A Hierarchical Model for Extreme Wind Speeds

Lee Fawcett[†] and David Walshaw

School of Mathematics & Statistics, Newcastle University, UK

[†]Email: lee.fawcett@ncl.ac.uk

Motivation

- A typical extreme value analysis is often carried out on the basis of over–simplistic inferential procedures, though the data being analysed may be structurally complex.
- We construct a hierarchical model which *exploits* such meteorological structure, enabling increased precision for inferences at individual sites; the Bayesian paradigm provides the most natural setting for this.
- There are other reasons for working within the Bayesian framework when analysing extremes: increased precision and a natural extension to prediction, for example.

Wind Speed Data

We develop a hierarchical model for extreme wind speeds observed over a region of central and northern England.

- Data were collected hourly, over a period of 18 years (1974–1991 inclusive).
- Figure 1 shows the location of the 12 wind speed stations, as well as exploratory analyses for two contrasting sites — Bradfield and Nottingham.
- Notice the site and seasonal variation present, as well as substantial short–term temporal dependence.

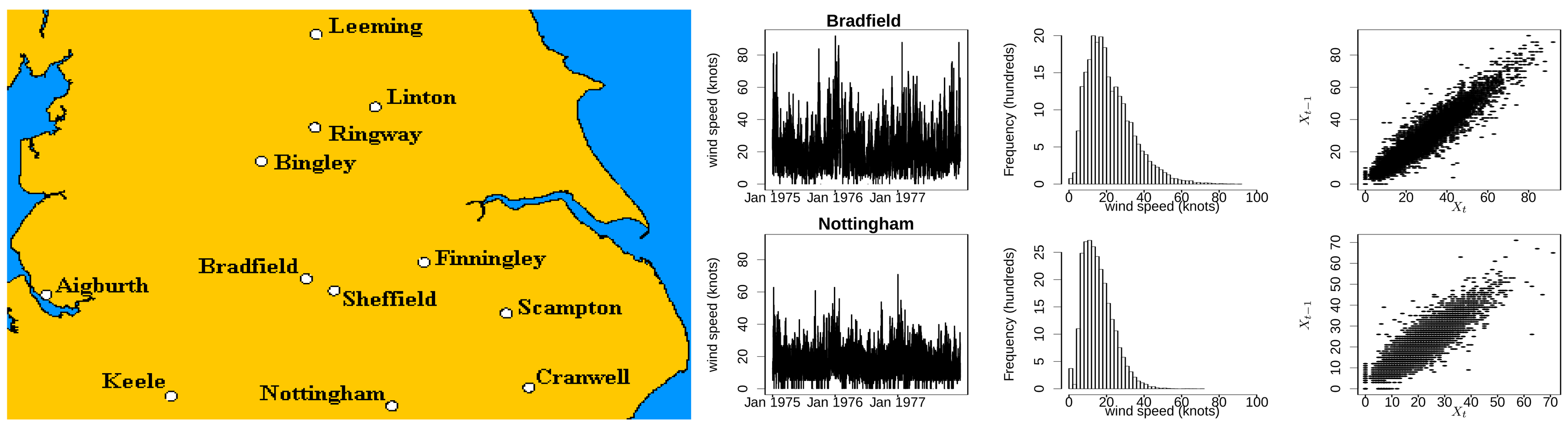

Figure 1: Location of wind speed stations and exploratory analyses

A Hierarchical Model

A model for extremes

Let $X_1, X_2, \ldots$ be a sequence of IID random variables with common distribution function F. For some high threshold u, the distribution of $(X - u|X > u)$ will be approximately of **Generalised Pareto Distribution** (GPD) form:

$$H(y) = 1 - \left(1 + \frac{\xi y}{\sigma}\right)_+^{-1/\xi}, \qquad (1)$$

where σ and ξ are scale and shape parameters (respectively). We reparameterise the scale parameter to $\tilde{\sigma} = \sigma - \xi u$ to avoid problems of threshold–dependency (Fawcett, 2005).

Site and seasonal variation

We take a pragmatic approach to site and seasonal variation, fitting a separate GPD to each season within each site. We use calendar months to define seasons, giving

$$(\tilde{\sigma}_{m,j}, \xi_{m,j}), \qquad m, j = 1, \ldots, 12,$$

where m and j are indices of season and site (respectively).

Temporal dependence

Preliminary analyses in Fawcett & Walshaw (2006a) suggest a first–order Markov structure for the serial dependence present in the wind speed extremes.

- Given a model $f(x_i, x_{i+1}; \boldsymbol{\psi})$ specified by parameter vector $\boldsymbol{\psi}$, the likelihood for $\boldsymbol{\psi}$ is:

$$L(\boldsymbol{\psi}) = f(x_1; \boldsymbol{\psi}) \prod_{i=1}^{n-1} f(x_i, x_{i+1}; \boldsymbol{\psi}) \Big/ \prod_{i=1}^{n-1} f(x_i; \boldsymbol{\psi}). \qquad (2)$$

- A bivariate extreme value distribution for contributions to the numerator in (2) is the **logistic model**, with:

$$F(x_i, x_{i+1}) = 1 - \left(Z(x_i)^{-1/\alpha} + Z(x_{i+1})^{-1/\alpha}\right)^{\alpha}, \qquad x_i, x_{i+1} > u,$$

where the transformation Z ensures that the margins are of GPD form (1). Independence and complete dependence are obtained when $\alpha = 1$ and $\alpha \searrow 0$.

Model specification

- We now build the following random effects model:

$$\log(\tilde{\sigma}_{m,j}) = \gamma_{\tilde{\sigma}}^{(m)} + \epsilon_{\tilde{\sigma}}^{(j)}, \qquad \xi_{m,j} = \gamma_{\xi}^{(m)} + \epsilon_{\xi}^{(j)} \qquad \text{and} \qquad \alpha_j = \epsilon_{\alpha}^{(j)},$$

where $\gamma_\bullet^{(m)}$ and $\epsilon_\bullet^{(j)}$ represent seasonal and site effects respectively.

- All random effects for $\log(\tilde{\sigma}_{m,j})$ and $\xi_{m,j}$ are taken to be Normally distributed. We adopt a conditional autoregressive structure for the seasonal effects, giving:

$$\gamma_\bullet^{(m)} | \left(\gamma_\bullet^{(m-1)}, \gamma_\bullet^{(m+1)}\right) \sim N\left(\frac{1}{2}\left\{\gamma_\bullet^{(m-1)} + \gamma_\bullet^{(m+1)}\right\}, \tau_\bullet^{-1}\right),$$

$$\epsilon_\bullet^{(j)} \sim N\left(a_\bullet, \zeta_\bullet^{-1}\right) \text{ and } \epsilon_\alpha^{(j)} \sim U(0, 1).$$

- To attain conjugacy, the final layer of the model becomes:

$$a_\bullet \sim N\left(b_\bullet, c_\bullet^{-1}\right), \qquad \tau_\bullet \sim Ga(d_\bullet, e_\bullet), \text{ and } \qquad \zeta_\bullet \sim Ga(f_\bullet, g_\bullet).$$

MCMC Simulations and Return Levels

Each component in the hierarchical model is updated singly using a Gibbs sampler where conjugacy allows; elsewhere, we adopt a Metropolis step.

Some results: Bradfield in January

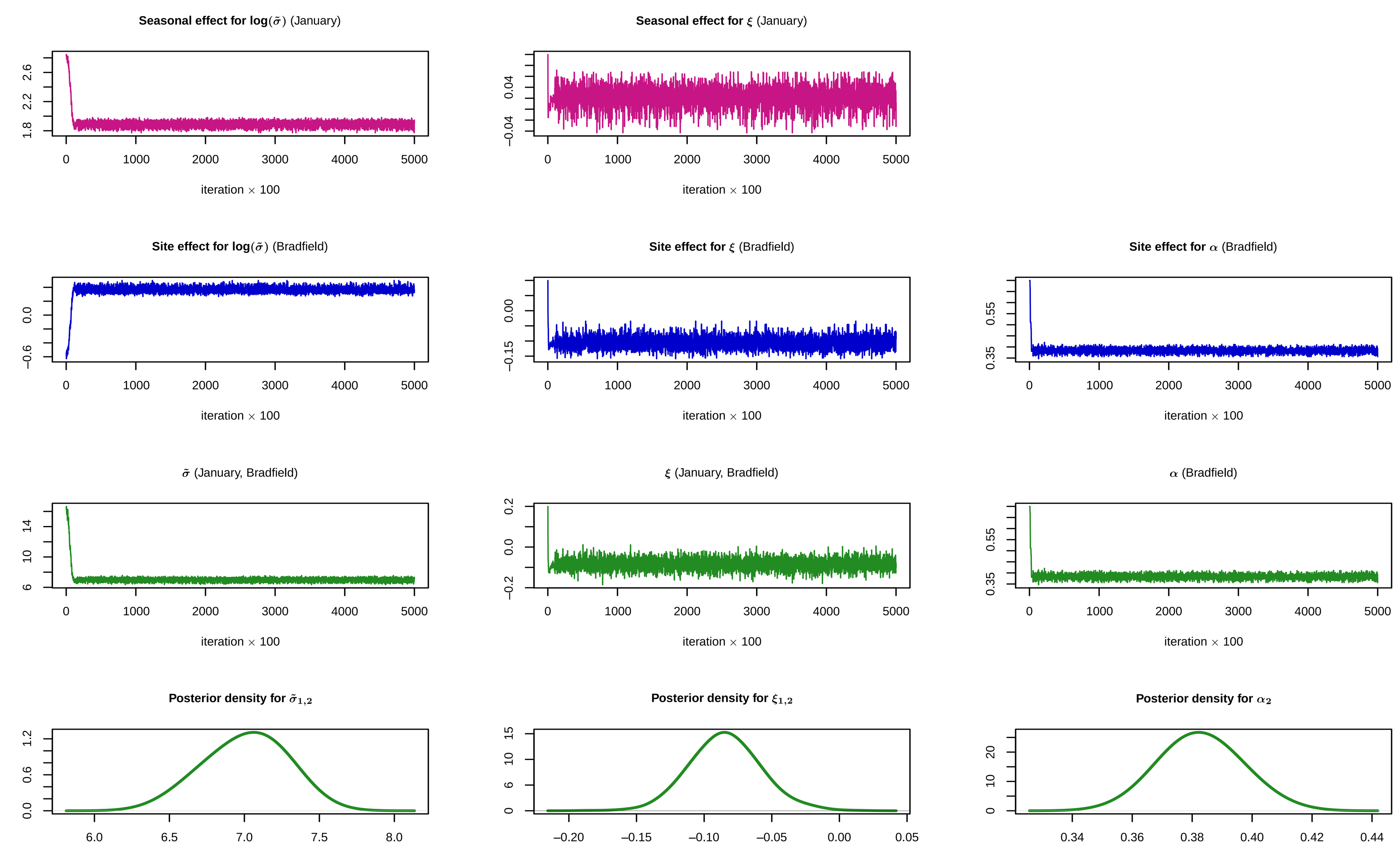

Figure 2: Markov chain Monte Carlo output for Bradfield in January

Return level inference

From a practical viewpoint, estimates of extreme quantiles – or **return levels** – are of greater interest. Such estimates are used as design specifications for buildings.

- For each site j, we consider the overall annual exceedance rate of some level q_r across all months using a GPD($\tilde{\sigma}_{m,j}$, $\xi_{m,j}$) and serial dependence parameter α_j.
- We set this equal to $1/r$ and solve numerically for q_r.
- q_r is then an estimate of the level which is exceeded once every r years at site j.
- Estimates of q_r are shown in Table 1, along with **predictive return level** estimates.

	Bradfield				Nottingham			
	q_{10}	q_{50}	q_{200}	q_{1000}	q_{10}	q_{50}	q_{200}	q_{1000}
Hierarchical model	96.9 (1.0)	103.5 (1.3)	112.5 (2.0)	128.1 (2.7)	66.5 (0.9)	73.5 (1.5)	79.5 (2.0)	86.3 (2.7)
Maximum likelihood	96.7 (2.9)	103.2 (5.9)	108.2 (8.8)	113.3 (12.2)	66.4 (2.6)	73.1 (4.9)	74.0 (11.2)	117.7 (14.6)
Predictive	*104.4*	*113.1*	*120.0*	*127.3*	*68.4*	*85.5*	*101.2*	*108.8*

Table 1: Return level estimates (knots) from the Bayesian hierarchical model (posterior mean and st. dev.) and a corresponding maximum likelihood fit (MLE and e.s.e.)

Summary and References

- Note a reduction in sampling variation under the Bayesian hierarchical model due to the pooling of information across sites and seasons.
- Maximum likelihood estimates of return levels can be very unstable, particularly for long–range return levels (see MLE of q_{1000} for Nottingham in Table 1).
- The Bayesian paradigm extends naturally to predictive return levels, wherein uncertainty in model estimation and future observations have been accounted for.
- Future work will investigate the assumption of spatial exchangeability and the elicitation of expert prior information.

Coles, S.G. (2001). A Random Effects Analysis of Extreme Wind Speed Data. *Preprint*. University of Padua, Padua.
Fawcett, L. (2005). Statistical Methodology for the Estimation of Environmental Extremes. PhD Thesis, Newcastle University.
Fawcett, L. and Walshaw, D. (2006). A Hierarchical Model for Extreme Wind Speeds. *Applied Statistics*, **55**, 5, 631—646.
Fawcett, L. and Walshaw, D. (2006a). Markov Chain Models for Extreme Wind Speeds. *Environmetrics*, **17**, 8, 795—809.